Motion capture enables precise, quantitative analysis of gesture and sign language production.

To introduce a wider audience of researchers to this field of inquiry, we present a review of previous works that utilized motion capture to study sign and gesture production along with comments on technical and methodological issues.

**Sign Languages**

While earlier motion capture studies on sign languages have focused on single signs or short sequences (Wilbur, 1990; Wilcox, 1992), more recent work deals with continuous, longer-duration, conversational data, and aims at supporting automated transcription and translation, as well as synthesis.

**Articulation and Prosody**

Hyperspectral data of American Sign Language (ASL) production has been studied extensively through motion capture. Signers typically use their hands, head, and body movements to convey meaning in ASL, and marker-based optical motion capture is well-suited for capturing these movements. More recently, work deals with continuous, longer-duration, conversational data, and aims at supporting automated transcription and translation, as well as synthesis.

**Synthesis**

Synthesis (or generation) pertains to sign production by animated (ID, cartoon, or robotic) characters. Lu & Hummel (2012) describe the design of an ASL database to support synthesis, as well as an evaluation of their design and recording methods. In later work, they used data from a multimodal hybrid motion system to train a vector-based language model to improve the understanding of feeding verbs. They report on their experiences in collecting and processing marker-based motion capture data for Field, research. They offer strategies for working with off-the-shelf setup, marker placement, and post-processing automation. Technical issues are important for developing a protocol for optimal data collection. A crucial challenge is to thoroughly investigate the methods to the hands of deaf individuals. Lu & Hummel (2012) present issues and other considerations for motion capture data collection. Wilcox et al. (2010) describe the creation of an annotated FSL database. Mehmet Aydın Baytaş (2012) report on their experiences in collecting and processing marker-based motion capture for Field, research. They offer strategies for working with off-the-shelf setup, marker placement, and post-processing automation. Technical issues are important for developing a protocol for optimal data collection. A crucial challenge is to thoroughly investigate the methods to the hands of deaf individuals.

**Dialogue**

Bosse et al. (2006) have used motion capture to study the computationalization of electric prompts and their effects on dialogue. Tyrone et al. (2010) have demonstrated the effect of communicative intent on the kinematics of functional gestures, while Masters et al. (2014) investigated how the form of pointing gestures can be affected by communicative intent. Data for all of these studies has been collected using optical motion capture, comprising 13 markers on the dominant hand only, during the data collection signers performed a sequence. Processes have been possible by leveraging the superhuman temporal and spatial resolution of marker-based optical motion capture.

**Technological Considerations and Limitations**

Motion capture studies on co-speech gestures often aim to understand gestures in everyday communication, language development, and speech/language impairments. More recent studies have also been motivated by topics in computing: activity and affect recognition, machine translation, natural avatar animation, and improving multimodal data analysis methods.

**Cost**

Marker-based and electromagnetic motion capture systems that afford high spatial and temporal resolution can be expensive. Markerless systems with a depth camera that detect human joint positions may not provide the resolution and accuracy that are required. Achieving high-resolution motion capture with the latter systems is an active research field (e.g., Amin et al., 2010).

**Noise**

Optical motion systems can be susceptible to interference from infrared light (e.g., sunlight) and reflective material in the environment. Other systems may require low light conditions.

**Accuracy and Precision**

While marker-based optical systems can make precise measurements during human-computer interaction, markerless systems can be used to capture pose, joint positions, or states of start-of-the-art tracking systems directly on the order of the resolution (Dhulak et al., 2017). Electromagnetic systems offer lower latency in terms of frame rate and can be used in real-time applications. Cameras can be used for both accuracy and precision and are appropriate for the inquiry at hand.

**Occlusions**

Optical marker systems require a clear line of sight between the camera and the markers or beacons identifying targets. Electric systems using reflective markers can mitigate occlusion issues. For this reason, while researchers studying co-speech gestures often use saved data, which can be common to understand that different software in the analysis pipeline can lead to different results and that different pipelines can also be used for recording data. The capture system must be selected in an informed fashion and confronted with pilot studies.

**Sampling Rate and Resolution**

A high sampling rate and resolution can capture even fast and remote movements in detail. The trade-off is that higher data rates will require more powerful hardware and computing resources, and only costly high-end motion capture will be able to capture each small movement. This is especially problematic for markerless systems that different software in the analysis pipeline can lead to different results and that different pipelines can also be used for recording data. The capture system must be selected in an informed fashion and confronted with pilot studies.

**Skill**

Collecting data with almost any motion capture system requires researchers to be knowledgeable about its technical limitations, as well as the specific goals of the study. The physical capabilities and limitations of the system environment (environment, algorithms, along with the operators, recording, and analysis of data. Conditioning sign and motion capture research, interpretability and learning in collaboration.

**References**


